The battle of neighborhoods or the battle of "*organic Food & Beverages*" at the close vicinity of Berlin metro stations

Final Report:

Capstone IBM Coursera Project for Data Science

Dr. B. Bayer, June 2021

Table of Content

List of Figures

1. Introduction

1.1 Background Information

Berlin is the capital of Germany with about 3.7 million inhabitants and is a city full of movement. On average, every Berliner travels more than three times a day. Local public transport plays a special role. Approx. 50 % of Berlin's households are car-free, and with 324 cars per 1,000 inhabitants, the city has the lowest motorization rate in Germany. In addition, even people who do have a car often use other means of transportation. Public transportation plays also a central role for commuting to work. For example, about 40% of all commuters use public transportation to reach their work destination. [1]

Another strongly growing trend is the demand for organic products and therefore incoming, the desire for a healthy diet [2]. In train stations and metro stations of Berlin, there are already numerous possibilities for food intake, and so numerous small and large stores offer food and drinks to take away. However, these offers are very often not considered healthy at all and thus contradict the desire for a healthy diet. The market for healthy take-away products to be bought on the way to work is still relatively small and this results in a large potential market for investors.

1.2 Problem Statement

With the aforementioned prospect, various stakeholders (entrepreneurs, investors) may be interested to explore the organic food and beverage shop business opportunities in the very close vicinity of Berlin subway system. This data science project is thus carried out to help them answer the following question: Which of Berlin metro stations are strategic for opening an "organic food and beverage" business?

2. Data

In order to explore potential answer to the problem, the following data were required.


https://de.wikipedia.org/wiki/Liste_der_Berliner_U-Bahnhöfe

Stationen [Bearbeiten | Quelltext bearbeiten]

Inhaltsverzeichnis A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Bahnhof (Kürzel) Karte	Linie	Eröffnung	Lage	Ortsteil	Umstieg	Denkmal	Anmerkungen	Sehenswürdigkeiten	Bild
Adenauerplatz (Ado) 52° 29′ 59″ N, 13° 18′ 26″ O	U7	28. Apr. 1978	Tunnel	Charlottenburg		–			
Afrikanische Straße (Afr) 52° 33′ 38″ N, 13° 20′ 3″ O	U6	3. Mai 1956	Tunnel	Wedding		–			
Alexanderplatz (A) 52° 31′ 17″ N, 13° 24′ 48″ O	U2	1. Juli 1913	Tunnel	Mitte		Eintrag in der Berliner Landesdenkmalliste		Alexa, Fernsehturm, Haus des Lehrers, Urania-Weltzeituhr	

Figure 1: Wikipedia information of Berlin tram stations

Geographical coordinates

Metro stations of Berlin city with their geographical coordinates were downloaded from Wikipedia website (see Figure 1 above):

https://de.wikipedia.org/wiki/Liste_der_Berliner_U-Bahnh%C3%B6fe [3]

They were required to utilize Foursquare API in the subsequent step to retrieve venue data.

Venue

"Foursquare is the most trusted, independent location data platform for understanding how people move through the real world [4]. Information about venues of the station: the names, category, venue latitudes, venue longitudes were obtained using the developer interface Foursquare API. The stations will be clustered based on their venues to find the best location candidates for "organic F&B". Find below Figure 2 for venue data as pandas data frame.

```
[38]: station_venues_all.head()
```

[38]:

	Station	Station Latitude	Station Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	Adenauerplatz	52.499722	13.307222	Bellucci	52.499430	13.306800	Italian Restaurant
1	Adenauerplatz	52.499722	13.307222	Block House	52.499645	13.306755	Steakhouse
2	Adenauerplatz	52.499722	13.307222	Adenauerplatz	52.499932	13.307214	Plaza
3	Adenauerplatz	52.499722	13.307222	Einstein Kaffee	52.500056	13.306058	Coffee Shop
4	Adenauerplatz	52.499722	13.307222	Rossmann	52.500013	13.308113	Drugstore

Figure 2: Foursquare venue data

3. Methodology

This section represents the main components of the report the chosen chapters follow the recommendation of coursera. It starts with data extraction (**web scraping**) of Berlin metro stations and retrieval of geographical coordinates. Leveraging Foursquare API, these coordinate data are given as inputs to explore venues of the stations.

One-hot encoding is performed to analyze and narrow down the most common venues in each of the station. Given all the venues surrounding them, the stations are clustered using K-means algorithm. The number of optimal **clusters** is decided using the elbow method and silhouette score. Each cluster is separately analyzed to examine one discriminating venue that characterizes them. Analysis of the clusters and visualization will give insights as to where the strategic regions to set up the business.

3.1 Web scraping (of tram stations)

A total number of 178 tram stations were taken into account and scraped from Wikipedia (refer to previous Chapter “data”). The coordinates were published in the format of degree including minutes and seconds. Before further automatic use within Foursquare, the coordinates were converted into decimal degree (Figure 3).

```
# Convert coordinates given in degree to dezimal
data['Latitude'] = ((( data['lat_sec'] / 60 ) +  data['lat_min']) / 60) + data['lat_grad']
data['Longitude'] = ((( data['long_sec'] / 60 ) +  data['long_min']) / 60) + data['long_grad']
```

Figure 3: Conversion from degree to decimal degree

The figures below show an example of the final pandas data frame (Figure 4) and a map of stations distribution (Figure 5) generated with python library folium. Please note the lack of tram stations in the eastern part of the city. This is due to the former separation of East and West Berlin. The eastern area still has high potential for development of the public transport system.

	Locality	Station	Latitude	Longitude
0	Charlottenburg	Adenauerplatz	52.499722	13.307222
1	Wedding	Afrikanische Straße	52.560556	13.334167
2	Mitte	Alexanderplatz	52.521389	13.413333
5	Spandau	Altstadt Spandau	52.539167	13.205556
6	Mariendorf	Alt-Mariendorf	52.439722	13.387500
...	...	...	...	...
194	Gropiusstadt	Wutzkyallee	52.423333	13.474722
195	Schöneberg	Yorckstraße	52.493056	13.370833
196	Haselhorst	Zitadelle	52.537778	13.217778
197	Charlottenburg	Zoologischer Garten	52.507222	13.332500
199	Gropiusstadt	Zwickauer Damm	52.423333	13.483889

178 rows × 4 columns

Figure 4: Geographical coordinates of the stations

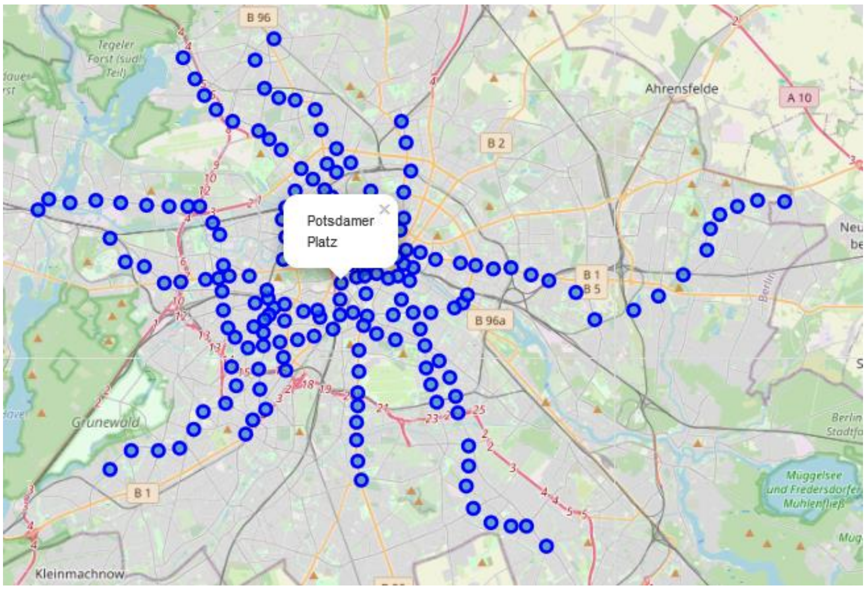

Figure 5: Map of the stations

3.2 Exploring venues / feature extracting

The search radius for each station was chosen quite narrow with 100 m to ensure the idea of buying food directly on the way to work. The API calls to Foursquare results in a total of 842 venues from 184 unique categories. The calls were made using a user-defined function. For the final data frame, categories which have nothing to do with food supply, such as the “Metro / Bus Station”, Gift Shop, “Bookstore”, etc, were removed from the data with subsequently 85 unique categories being considered. The Figure 6 shows the removal list.

```
[41]: removal_list = ['Drugstore', 'Bus Stop', 'Metro Station', 'Gift Shop', 'Clothing Store', 'Bookstore', 'Cosmetics Shop',
        'Department Store', 'Electronics Store', 'Shoe Store','Neighborhood', 'Mobile Phone Shop', 'Movie Theater','Supermarket',
        'Arts & Crafts Store', 'Liquor Store','Paper / Office Supplies Store', 'Bank', 'Grocery Store', 'Gym / Fitness Center', 'Flower Shop',
        'Organic Grocery', 'Vacation Rental', 'Athletics & Sports', 'Indie Movie Theater', 'Spa', 'Art Gallery', 'Museum', 'Pharmacy',
        'Big Box Store', 'Park', 'Farm', 'Garden Center', 'Gym','Opera House', 'Sporting Goods Shop', 'IT Services',
        'Furniture / Home Store', 'Hobby Shop', 'Roof Deck','Health & Beauty Service', 'History Museum','Photography Studio', 'Playground',
        'Pet Store', 'Theater', 'Public Bathroom','Flea Market', 'Toy / Game Store', 'Thrift / Vintage Store', 'Tram Station',
        'Mini Golf', 'Chocolate Shop', 'Optical Shop', 'Boutique',"Men's Store", 'Lake', 'Shipping Store',
        'Historic Site', 'Music Venue', 'Medical Center', 'Camera Store''Science Museum', 'Souvenir Shop', 'ATM',
        'Record Shop', 'Convenience Store','Automotive Shop', 'Miscellaneous Shop', 'Boarding House',
        'Cafeteria', 'Farmers Market', 'Hotel','Train Station', 'Salad Place', 'Newsstand','Smoke Shop', 'Taxi Stand', 'Lingerie Store',
        'Hot Spring', 'Nightclub', 'Beer Garden', 'Hostel','Yoga Studio', 'Light Rail Station', 'Performing Arts Venue',
        'Soccer Field', 'Beach Bar', 'Event Space', 'Multiplex','Platform', 'Post Office', 'Indie Theater', 'Discount Store',
        'Luggage Store', 'Escape Room', 'Coworking Space','Gas Station', 'Music Store', 'Lounge','Monument / Landmark', 'Plaza',
        "Women's Store", 'Speakeasy', 'Adult Boutique','Rental Car Location','Campground','Dance Studio', 'Comedy Club', 'Massage Studio',
        'Other Repair Shop', 'Bath House', 'Deli / Bodega', 'Waterfront','Garden','Trail','Martial Arts School',
        'Gay Bar', 'Hot Dog Joint','Circus', 'Gym Pool', 'Bridge', 'Rock Club', 'Stationery Store',
        'Pool', 'Business Service', 'Shopping Plaza','Hardware Store', 'Laundromat', 'Building','Scenic Lookout', 'Country Dance Club',
        'Concert Hall', 'Fountain', 'Wings Joint', 'Sculpture Garden', 'Plaza', 'Camera Store', 'Bubble Tea Shop',
        'Costume Shop',  'Video Store','Outdoor Supply Store', 'Bike Shop', 'General Entertainment', 'Art Museum', 'Sauna / Steam Room',
        'Bridal Shop','Print Shop', 'Intersection', 'Zoo Exhibit','Harbor / Marina', 'Pool Hall', 'Exhibit', 'Perfume Shop', 'Karaoke Bar',
         'Salon / Barbershop','Memorial Site', 'Jewelry Store', 'Carpet Store', 'Bike Rental / Bike Share']
```

Figure 6: Removal list for venue categories

The next two Figure 7 and Figure 8 show the frequency of each 85 categories and the most 20 frequented categories. 84 bakeries, 30 Doner restaurants (Turkish), 29 Cafés, 26 Coffee Shops and 24 Italian restaurants are the top 5 categories.

Please note that “vegetarian / vegan food places” indicating “healty” food are not at all among the top places instead found on place 18.

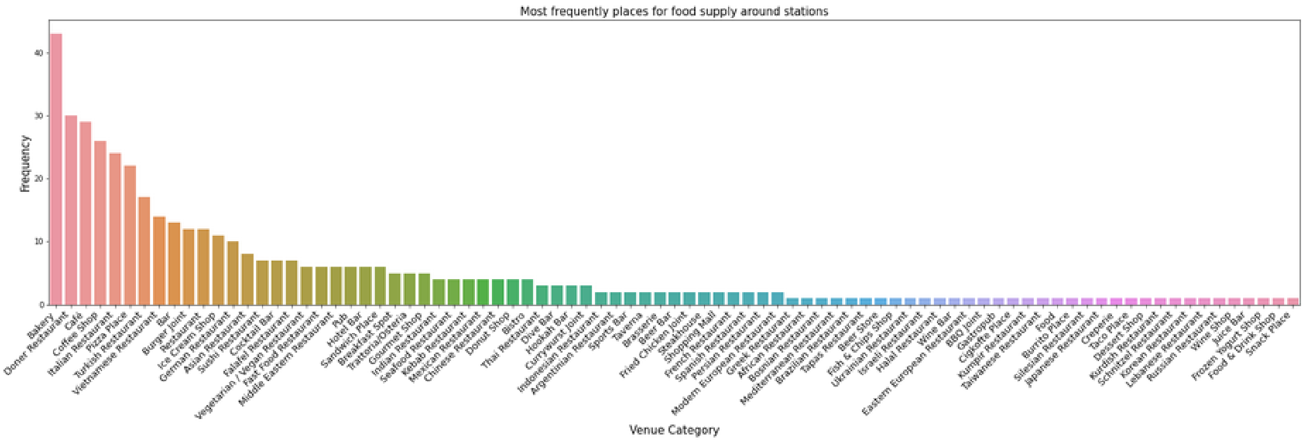

Figure 7: Frequency of all venue categories

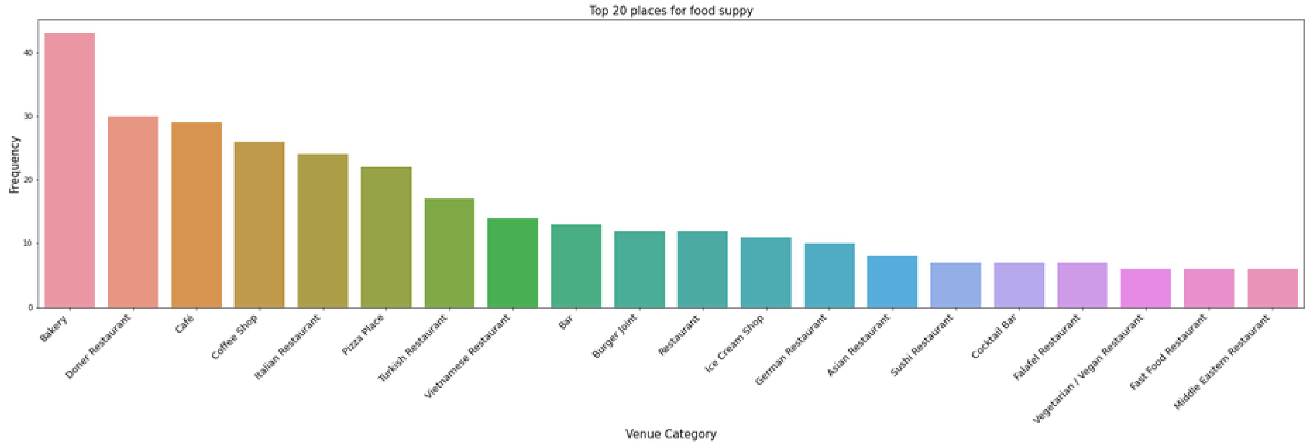

Figure 8: Frequency top 20 venue categories

3.3 One-hot encoding

One-hot encoding converts categorical variables (i.e., venues) into numeric variables. In this case, a dummy of all the venues was made and the mean of the frequency of venue occurrence were calculated. The data frame is then grouped by stations, as shown in Figure 9 below.

[50]:
```
station_grouped = station_onehot.groupby('Station').mean().reset_index()
print(station_grouped.shape)
station_grouped.head()
```

(133, 86)

[50]:

	Station	African Restaurant	Argentinian Restaurant	Asian Restaurant	BBQ Joint	Bakery	Bar	Beer Bar	Beer Store	Bistro	...	Tapas Restaurant	Taverna	Thai Restaurant	Trattoria/Osteria	Turkish Restaurant	Ukrainian Restaurant	Vegetarian / Vegan Restaurant	Vietnamese Restaurant	Wine Bar	Wine Shop
0	Adenauerplatz	0.0	0.000000	0.0	0.0	0.142857	0.0	0.0	0.0	0.142857	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
1	Afrikanische Straße	0.0	0.333333	0.0	0.0	0.333333	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
2	Alexanderplatz	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
3	Alt-Mariendorf	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	Alt-Tegel	0.0	0.000000	0.0	0.0	0.000000	0.0	0.0	0.0	0.000000	...	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

5 rows × 86 columns

Figure 9: One-hot encoding

The data were then filtered with a user-defined function to obtain five most common venues in each station, as displayed as a matrix in Figure 10.

```
[53]: venues_sorted_station.head(15)
```

[53]:

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue
0	Adenauerplatz	Steakhouse	Coffee Shop	Cocktail Bar	Bakery	Italian Restaurant
1	Afrikanische Straße	Hookah Bar	Argentinian Restaurant	Bakery	Food & Drink Shop	Dive Bar
2	Alexanderplatz	Coffee Shop	Doner Restaurant	Burger Joint	Wine Shop	Currywurst Joint
3	Alt-Mariendorf	Greek Restaurant	Italian Restaurant	Wine Shop	Food	Dive Bar
4	Alt-Tegel	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar
5	Alt-Tempelhof	Vietnamese Restaurant	Doner Restaurant	Fried Chicken Joint	Wine Shop	Food
6	Altstadt Spandau	Cocktail Bar	Bakery	Wine Shop	Food	Dive Bar
7	Amrumer Straße	Coffee Shop	Vietnamese Restaurant	Food	Dessert Shop	Dive Bar
8	Augsburger Straße	Turkish Restaurant	Café	Spanish Restaurant	Wine Shop	Fish & Chips Shop
9	Bayerischer Platz	Wine Shop	Bakery	German Restaurant	Currywurst Joint	Coffee Shop
10	Berliner Straße	Argentinian Restaurant	Chinese Restaurant	Wine Shop	Food	Dive Bar
11	Bernauer Straße	Coffee Shop	Vegetarian / Vegan Restaurant	Doner Restaurant	Bar	Food & Drink Shop
12	Birkenstraße	Café	Vegetarian / Vegan Restaurant	German Restaurant	Doner Restaurant	Restaurant
13	Bismarckstraße	Asian Restaurant	Café	Wine Shop	Food	Dive Bar
14	Blissestraße	Vietnamese Restaurant	Doner Restaurant	Bakery	Burger Joint	Sushi Restaurant

Figure 10: Top 5 common venue for each station

3.4 Unsupervised Machine Learning (Clustering)

The stations were clustered based on a set of similar characteristics or features, i.e., their surrounding venues. *K*-Means clustering, which was used in this part of the analysis, is an unsupervised machine learning algorithm that creates homogeneous subgroups/clusters from unlabeled data such that data points in each cluster are as similar as possible to each other according to a similarity measure (e.g., Euclidian distance).

A value of *k* (number of clusters) needs to be defined before proceeding with the clustering. The "*Elbow Method*" was used, which calculates the sum of squared distances of data points to their closest centroid (cluster center) for different values of *k*. The optimal value of *k* is the one after which there is a plateau (no significant decrease in sum of squared distances).

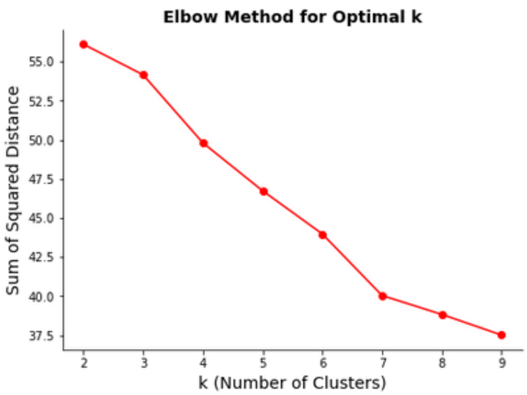

Figure 11: Elbow methode

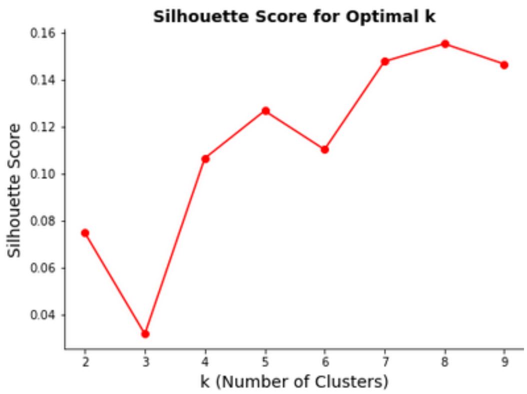

Figure 12: Silhouette methode

Because there is no discernible "*elbow*" from the plot (Figure 11) another measure was used: "*Silhouette Score*". Silhouette score varies from -1 to 1. A score value of 1 means the cluster is dense and well-separated from other clusters. A value nearing 0 represents overlapping clusters, data points are close to the decision boundary of neighboring clusters. A negative score indicates that the samples might have been assigned into the wrong clusters. Given that there is a peak at $k = 5$ (Figure 12), the *K*-Means clustering was proceeded with that value.

However, both methods, the elbow and silhouette, are not very clearly and need further investigation.

4. Results

As a result of the clustering, a merged data frame with top five most common categories for each station and attributed to one of the five cluster group (0-4) can be given (Figure 13). The clusters were color-coded and visualized on a map of Berlin (Figure 14) to understand how they are distributed across the city.

```
[63]: data_station_cluster_merged = data_stat
      # merge top venues_sorted with Berlin data_stat from the beginning
      data_station_cluster_merged = venues_sorted_station.join(data_station_cluster_merged.set_index('Station'), on='Station')
      data_station_cluster_merged.head(50) # check the last columns
```

[63]:

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
0	Adenauerplatz	Steakhouse	Coffee Shop	Cocktail Bar	Bakery	Italian Restaurant	3	Charlottenburg	52.499722	13.307222
1	Afrikanische Straße	Hookah Bar	Argentinian Restaurant	Bakery	Food & Drink Shop	Dive Bar	3	Wedding	52.560556	13.334167
2	Alexanderplatz	Coffee Shop	Doner Restaurant	Burger Joint	Wine Shop	Currywurst Joint	4	Mitte	52.521389	13.413333
3	Alt-Mariendorf	Greek Restaurant	Italian Restaurant	Wine Shop	Food	Dive Bar	2	Mariendorf	52.439722	13.387500
4	Alt-Tegel	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Tegel	52.589444	13.283611
5	Alt-Tempelhof	Vietnamese Restaurant	Doner Restaurant	Fried Chicken Joint	Wine Shop	Food	4	Tempelhof	52.466111	13.385556
6	Altstadt Spandau	Cocktail Bar	Bakery	Wine Shop	Food	Dive Bar	0	Spandau	52.539167	13.205556
7	Amrumer Straße	Coffee Shop	Vietnamese Restaurant	Food	Dessert Shop	Dive Bar	3	Wedding	52.542222	13.348889
8	Augsburger Straße	Turkish Restaurant	Café	Spanish Restaurant	Wine Shop	Fish & Chips Shop	3	Charlottenburg	52.500556	13.336389
9	Bayerischer Platz	Wine Shop	Bakery	German Restaurant	Currywurst Joint	Coffee Shop	3	Schöneberg	52.488611	13.340000
10	Berliner Straße	Argentinian Restaurant	Chinese Restaurant	Wine Shop	Food	Dive Bar	3	Wilmersdorf	52.487222	13.330833

Figure 13: Merged dataframe with cluster labels for each station

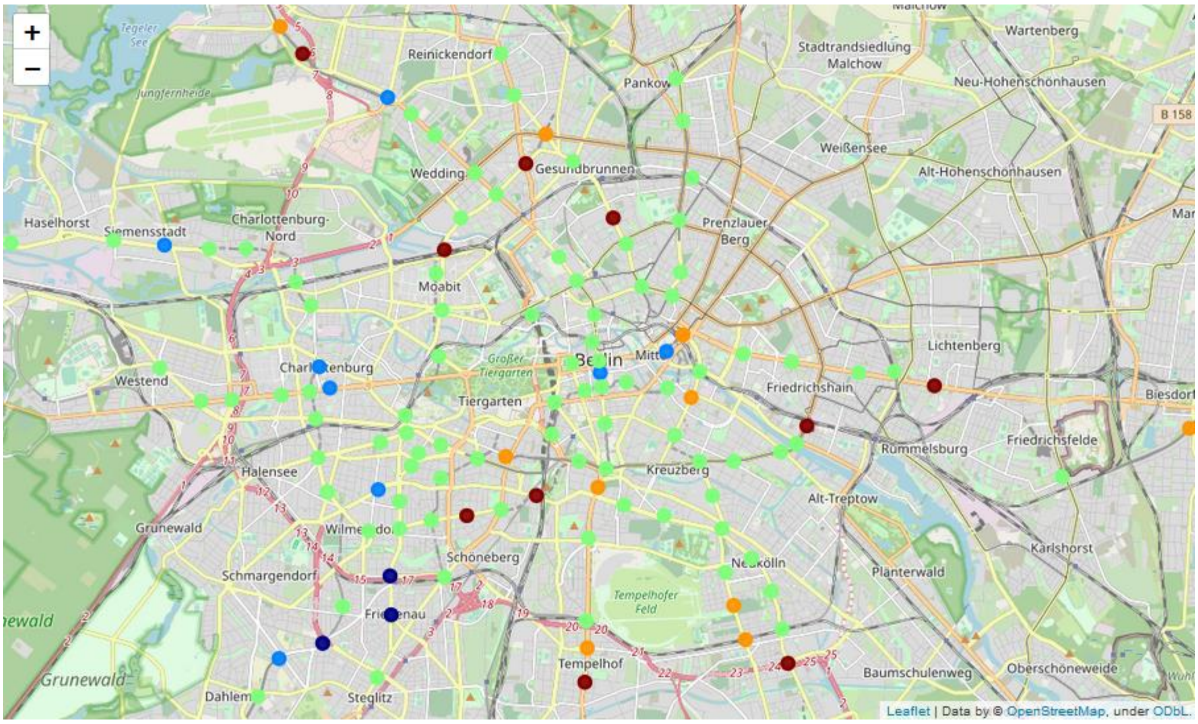

Figure 14: Clusters color coded

Find below an introduction of the findings for each cluster.

Cluster 0

(Color code in map: wine red (or brown for some eyes)

A total of 13 stations fall into this cluster with the following 1st and 2nd most common venue (Figure 15). First most common venue reflects “Bakeries” (Figure 16).

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
6	Altstadt Spandau	Cocktail Bar	Bakery	Wine Shop	Food	Dive Bar	0	Spandau	52.539167	13.205556
19	Britz-Süd	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	0	Britz	52.437778	13.448333
24	Eisenacher Straße	Vietnamese Restaurant	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	0	Schöneberg	52.489444	13.350278
32	Grenzallee	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	0	Britz	52.463333	13.443889
50	Kaiserin-Augusta-Straße	Bakery	Coffee Shop	Wine Shop	Food	Dive Bar	0	Tempelhof	52.460000	13.384722
65	Magdalenenstraße	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	0	Lichtenberg	52.512500	13.486389
74	Nauener Platz	Doner Restaurant	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	0	Wedding	52.551667	13.367500
84	Otisstraße	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	0	Reinickendorf	52.571111	13.302778
123	Voltastraße	Bakery	Bar	Gastropub	Wine Shop	Food	0	Gesundbrunnen	52.542222	13.393056
124	Warschauer Straße	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	0	Friedrichshain	52.505278	13.449167
126	Westhafen	Coffee Shop	Bakery	Wine Shop	Food	Dive Bar	0	Moabit	52.536389	13.343889
128	Wittenau	Doner Restaurant	Bakery	Wine Shop	Food & Drink Shop	Dive Bar	0	Wittenau	52.595833	13.336667
130	Yorckstraße	Dive Bar	Bakery	Wine Shop	Food & Drink Shop	Doner Restaurant	0	Schöneberg	52.493056	13.370833

Figure 15: Cluster 0 top common venues

	1st Most Common Venue	Count
0	Bakery	7
1	Doner Restaurant	2
2	Vietnamese Restaurant	1
3	Cocktail Bar	1
4	Coffee Shop	1
5	Dive Bar	1

	2st Most Common Venue	Count
0	Bakery	6
1	Wine Shop	5
2	Coffee Shop	1
3	Bar	1

Figure 16: Cluster 0 - 1st and 2nd common venue

Cluster 1

(Color code in map: dark blue)

With three members (stations) the smallest cluster with prominent venue restaurants (second most uniform Mexican, **Figure 17**).

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
18	Breitenbachplatz	Dessert Shop	Mexican Restaurant	Food	Dive Bar	Doner Restaurant	1	Dahlem	52.466944	13.308611
20	Bundesplatz	Middle Eastern Restaurant	Mexican Restaurant	Wine Shop	Fish & Chips Shop	Dive Bar	1	Wilmersdorf	52.478889	13.328056
29	Friedrich-Wilhelm-Platz	Spanish Restaurant	Mexican Restaurant	Wine Shop	Creperie	Dessert Shop	1	Friedenau	52.471944	13.328611

Figure 17: Cluster 1 top common venues

Cluster 2

(Color code in map: brighter blue)

A total sum of thirteen stations fall into this cluster (Figure 18).
Italian restaurants and Wine Shops are prominent features (Figure 19).

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
3	Alt-Mariendorf	Greek Restaurant	Italian Restaurant	Wine Shop	Food	Dive Bar	2	Mariendorf	52.439722	13.387500
16	Borsigwerke	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	2	Tegel	52.581944	13.290833
22	Deutsche Oper	Italian Restaurant	Chinese Restaurant	Wine Shop	Food	Dive Bar	2	Charlottenburg	52.511944	13.310556
28	Französische Straße	Gourmet Shop	Restaurant	Italian Restaurant	Fish & Chips Shop	Dessert Shop	2	Mitte	52.514722	13.389167
43	Hohenzollernplatz	Restaurant	Food	Dessert Shop	Dive Bar	Doner Restaurant	2	Wilmersdorf	52.494167	13.324722
61	Kurt-Schumacher-Platz	Restaurant	Doner Restaurant	Italian Restaurant	Bistro	Food	2	Reinickendorf	52.563333	13.327500
89	Podbielskiallee	Restaurant	Food	Dessert Shop	Dive Bar	Doner Restaurant	2	Dahlem	52.464167	13.295833
95	Richard-Wagner-Platz	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	2	Charlottenburg	52.515833	13.307500
96	Rohrdamm	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	2	Siemensstadt	52.537222	13.262500
99	Rotes Rathaus	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	2	Mitte	52.518611	13.408333
100	Rudow	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	Doner Restaurant	2	Rudow	52.416111	13.495278

Figure 18: Cluster 2 top common venues

	1st Most Common Venue	Count
0	Italian Restaurant	6
1	Restaurant	3
2	Gourmet Shop	1
3	Greek Restaurant	1

	2st Most Common Venue	Count
0	Wine Shop	5
1	Food	2
2	Doner Restaurant	1
3	Chinese Restaurant	1
4	Italian Restaurant	1
5	Restaurant	1

Figure 19: Cluster 2 - 1st and 2nd common venue

Cluster 3

(Color code in map: bright green)

This cluster is by far the largest one with 92 member stations (Figure 20). Prominent are Coffee and Cafe places, followed by pizza and Turkish food. Wine Shops as 2nd favorite not to be forgotten (Figure 21 and Figure 22).

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
0	Adenauerplatz	Steakhouse	Coffee Shop	Cocktail Bar	Bakery	Italian Restaurant	3	Charlottenburg	52.499722	13.307222
1	Afrikanische Straße	Hookah Bar	Argentinian Restaurant	Bakery	Food & Drink Shop	Dive Bar	3	Wedding	52.560556	13.334167
7	Amrumer Straße	Coffee Shop	Vietnamese Restaurant	Food	Dessert Shop	Dive Bar	3	Wedding	52.542222	13.348889
8	Augsburger Straße	Turkish Restaurant	Café	Spanish Restaurant	Wine Shop	Fish & Chips Shop	3	Charlottenburg	52.500556	13.336389
9	Bayerischer Platz	Wine Shop	Bakery	German Restaurant	Currywurst Joint	Coffee Shop	3	Schöneberg	52.488611	13.340000
...	...	...	...	...	...	...	...	...	...	...
125	Weberwiese	Vietnamese Restaurant	Bar	Wine Shop	Food & Drink Shop	Dive Bar	3	Friedrichshain	52.516667	13.445000
127	Wilmersdorfer Straße	Pizza Place	Doner Restaurant	Bakery	Italian Restaurant	Café	3	Charlottenburg	52.506667	13.306667
129	Wittenbergplatz	Gourmet Shop	Turkish Restaurant	Burrito Place	Wine Shop	Food	3	Schöneberg	52.501944	13.343056
131	Zitadelle	Shopping Mall	Fast Food Restaurant	Wine Shop	Creperie	Dessert Shop	3	Haselhorst	52.537778	13.217778
132	Zoologischer Garten	Fried Chicken Joint	Wine Shop	Food	Dessert Shop	Dive Bar	3	Charlottenburg	52.507222	13.332500

92 rows × 10 columns

Figure 20: Cluster 3 top common venues

	1st Most Common Venue	Count
0	Coffee Shop	10
1	Café	9
2	Pizza Place	8
3	Ice Cream Shop	6
4	Turkish Restaurant	6
5	Hotel Bar	5
6	Vietnamese Restaurant	4
7	Asian Restaurant	4
8	German Restaurant	3
9	Pub	3
10	Hookah Bar	2
11	Italian Restaurant	2
12	Sandwich Place	2
13	Bakery	2
14	Wine Shop	1
15	Food	1
16	Kebab Restaurant	1
17	Food & Drink Shop	1
18	Steakhouse	1
19	Shopping Mall	1
20	Argentinian Restaurant	1
21	Indian Restaurant	1
22	BBQ Joint	1
23	Sushi Restaurant	1
24	Seafood Restaurant	1
25	Schnitzel Restaurant	1
26	Brazilian Restaurant	1
27	Burger Joint	1
28	Falafel Restaurant	1
29	Sports Bar	1
30	Brasserie	1
31	Cigkofte Place	1
32	Fried Chicken Joint	1
33	Persian Restaurant	1
34	Doner Restaurant	1
35	Gourmet Shop	1
36	Restaurant	1
37	Chinese Restaurant	1
38	Cocktail Bar	1
39	Halal Restaurant	1

Figure 21: Cluster 3 - 1st common venue

	2st Most Common Venue	Count
0	Wine Shop	14
1	Vietnamese Restaurant	6
2	Bakery	6
3	Pizza Place	4
4	Trattoria/Osteria	4
5	Coffee Shop	4
6	Café	4
7	Doner Restaurant	3
8	Turkish Restaurant	3
9	Vegetarian / Vegan Restaurant	3
10	Donut Shop	2
11	Bar	2
12	Fast Food Restaurant	2
13	German Restaurant	2
14	Dive Bar	2
15	Middle Eastern Restaurant	2
16	Restaurant	2
17	Indonesian Restaurant	2
18	Cocktail Bar	2
19	Italian Restaurant	2
20	Ice Cream Shop	2
21	Argentinian Restaurant	1
22	Israeli Restaurant	1
23	African Restaurant	1
24	Eastern European Restaurant	1
25	Chinese Restaurant	1
26	Kebab Restaurant	1
27	Wine Bar	1
28	Sports Bar	1
29	Pub	1
30	Ukrainian Restaurant	1
31	Sushi Restaurant	1
32	Hookah Bar	1
33	Persian Restaurant	1
34	Hotel Bar	1
35	Taverna	1
36	Thai Restaurant	1
37	Sandwich Place	1
38	Korean Restaurant	1
39	Food	1

Figure 22: Cluster 3 – 2nd common venue

Cluster 4

(Color code in map: orange)

This cluster results in fourteen member stations (Figure 23). Top venue is Doner restaurants (Figure 24).

	Station	1st Most Common Venue	2nd Most Common Venue	3rd Most Common Venue	4th Most Common Venue	5th Most Common Venue	Cluster_Labels	Locality	Latitude	Longitude
2	Alexanderplatz	Coffee Shop	Doner Restaurant	Burger Joint	Wine Shop	Currywurst Joint	4	Mitte	52.521389	13.413333
4	Alt-Tegel	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Tegel	52.589444	13.283611
5	Alt-Tempelhof	Vietnamese Restaurant	Doner Restaurant	Fried Chicken Joint	Wine Shop	Food	4	Tempelhof	52.466111	13.385556
25	Elsterwerdaer Platz	Asian Restaurant	Doner Restaurant	Italian Restaurant	Wine Shop	Food & Drink Shop	4	Biesdorf	52.505000	13.560556
40	Heinrich-Heine-Straße	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Mitte	52.510278	13.415833
42	Hermannstraße	Doner Restaurant	Ice Cream Shop	Breakfast Spot	Food	Dive Bar	4	Neukölln	52.467500	13.431389
44	Holzhauser Straße	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Tegel	52.575833	13.296111
52	Kaulsdorf-Nord	Doner Restaurant	Turkish Restaurant	Wine Shop	Food	Dessert Shop	4	Hellersdorf	52.521111	13.588889
58	Krumme Lanke	Doner Restaurant	Italian Restaurant	Wine Shop	Food & Drink Shop	Dive Bar	4	Zehlendorf	52.443333	13.241389
60	Kurfürstenstraße	Asian Restaurant	Doner Restaurant	Bakery	Wine Shop	Food & Drink Shop	4	Tiergarten	52.500000	13.361944
62	Leinestraße	Doner Restaurant	Bar	Bosnian Restaurant	Wine Shop	Food & Drink Shop	4	Neukölln	52.473611	13.428056
66	Mehringdamm	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Kreuzberg	52.494444	13.388611
81	Oskar-Helene-Heim	Doner Restaurant	Wine Shop	Food	Dessert Shop	Dive Bar	4	Dahlem	52.450278	13.269722
82	Osloer Straße	Doner Restaurant	Pizza Place	Bakery	Fast Food Restaurant	Wine Shop	4	Gesundbrunnen	52.556944	13.373333

Figure 23: Cluster 4 top common venues

	1st Most Common Venue	Count
0	Doner Restaurant	10
1	Asian Restaurant	2
2	Vietnamese Restaurant	1
3	Coffee Shop	1

	2st Most Common Venue	Count
0	Wine Shop	5
1	Doner Restaurant	4
2	Turkish Restaurant	1
3	Ice Cream Shop	1
4	Italian Restaurant	1
5	Bar	1
6	Pizza Place	1

Figure 24: Cluster 4 - 1st and 2nd common venue

5. Discussion

Exploratory data analysis as well as machine learning and visualization techniques have provided some insights into the problem at hand. A total of 842 items originated by 184 venue categories for all 178 Berlin metro stations regions were returned at the time the API call was made. The search radius was chosen quite narrow with 100 m as only food intake places (incl. to-go markets) are of interest for this study (according the idea of “something healthy to buy on the way to the office …”).

After removing venue categories not of interest for the regarded food industry 85 unique categories were being left and seen as data basis for the study. The most common categories overall are 1. Bakeries, 2. Doner restaurants, 3. Cafes, 4. Coffee shops, and 5. Italian and Pizza places (refer to Figure 8)

After deciding on an optimal k value of 5, K-Means algorithm was run to cluster the stations based on their most common surrounding venues. To determine this optimal k value, two common methods elbow and silhouette were applied. The result of k = 5 result is ambiguous and needs further discussions.

Each of the five clusters, labeled 0-4, is characterized by dominant venues as follows. A considerable number of coffee shops and bakeries as well as Turkish food and Wine Shops are present for the largest cluster 3.

Cluster Label	Member	Common Venue
0	13	Bakeries
1	3	Mexican and Wine Shops
2	13	Italian and Wine Shops
3	92	Coffee/Cafe, Pizza, Turkish food and Wine Shops
4	14	Doner restaurants and Wine Shops

Figure 25: Dominant venues overview for all clusters

Categories indicating "healthy" food, such as "vegan/vegetarian restaurants", are not rated as a top category. In fact, such eateries are very, very rare, and therefore it is recommended that stakeholders look for opportunities to open organic food and beverage stores throughout

Berlin's metro stations. As an alternative, already existing Bakeries and Café could focus on offering such products.

6. Conclusion

Stakeholders searching for opportunities to open organic food and beverages (incl. vegan / vegetarian dishes) may want to consider setting up their business someplace where competitions are not severe. This study has shown that in the very close proximity of metro/tram stations of Berlin (radius of 100 m) such places don't exist and, therefore, such places are among the best candidates for organic food and beverages location.

7. References

[1] https://www.cnb-online.de/hintergruende/zahlen-und-fakten-zum-oepnv/
[2] https://www.oekolandbau.de/handel/marktinformationen/europaeischer-bio-markt-waechst-auf-ueber-40-milliarden-euro/
[3] https://de.wikipedia.org/wiki/Liste_der_Berliner_U-Bahnh%C3%B6fe
[4] https://developer.foursquare.com/